

# The number of distinct and repeated squares and cubes in the Fibonacci sequence

Huang Yuke<sup>1</sup> Wen Zhiying<sup>2</sup>

## ABSTRACT

The Fibonacci sequence  $\mathbb{F}$  is the fixed point beginning with  $a$  of morphism  $\sigma(a, b) = (ab, a)$ . In this paper, we get the explicit expressions of all squares and cubes, then we determine the number of distinct squares and cubes in  $\mathbb{F}[1, n]$  for all  $n$ , where  $\mathbb{F}[1, n]$  is the prefix of  $\mathbb{F}$  of length  $n$ . By establishing and discussing the recursive structure of squares and cubes, we give algorithms for counting the number of repeated squares and cubes in  $\mathbb{F}[1, n]$  for all  $n$ , and get explicit expressions for some special  $n$  such as  $n = f_m$  (the Fibonacci number) etc., which including some known results such as in A.S.Fraenkel and J.Simpson[8, 9], J.Shallit et al[7].

**Key words:** the Fibonacci sequence; square; cube; algorithm; the sequence of return words.

## 1 Introduction

Let  $\mathcal{A} = \{a, b\}$  be a binary alphabet. The concatenation of factors  $\nu$  and  $\omega$  denoted by  $\nu\omega$ . The Fibonacci sequence  $\mathbb{F}$  is the fixed point beginning with  $a$  of the Fibonacci morphism  $\sigma$  defined by  $\sigma(a) = ab$  and  $\sigma(b) = a$ . As a classical example over a binary alphabet,  $\mathbb{F}$  having many remarkable properties, we refer to M.Lothaire[14, 15], J.M.Allouche and J.Shallit[1], Berstel[2, 3].

Let  $\omega$  be a factor of  $\mathbb{F}$ , denoted by  $\omega \prec \mathbb{F}$ . Since  $\mathbb{F}$  is uniformly recurrent,  $\omega$  occurs infinitely many times. Let  $\omega_p$  be the  $p$ -th occurrence of  $\omega$ . If the factor  $\omega$  and integer  $p$  such that  $\omega_p\omega_{p+1}$  (resp.  $\omega_p\omega_{p+1}\omega_{p+2}$ ) is the factor of  $\mathbb{F}$ , we call it a square (resp. cube) of  $\mathbb{F}$ . As we know,  $\mathbb{F}$  contains no fourth powers. The properties of squares and cubes are objects of a great interest in many aspects of mathematics and computer science etc.

We denote  $F_m = \sigma^m(a)$  for  $m \geq 0$  and define  $F_{-1} = b$ ,  $F_{-2} = \varepsilon$  (empty word). Define  $f_m = |F_m|$  the  $m$ -th Fibonacci number,  $f_{-2} = 0$ ,  $f_{-1} = 1$ ,  $f_{m+1} = f_m + f_{m-1}$  for  $m \geq -1$ . Let  $\mathbb{F}[1, n]$  be the prefix of  $\mathbb{F}$  of length  $n$ . In this paper, we consider the four functions below:

- $A(n) := \#\{\omega : \omega\omega \prec \mathbb{F}[1, n]\}$ , the number of distinct squares in  $\mathbb{F}[1, n]$ ;
- $B(n) := \#\{(\omega, p) : \omega_p\omega_{p+1} \prec \mathbb{F}[1, n]\}$ , the number of repeated squares in  $\mathbb{F}[1, n]$ ;
- $C(n) := \#\{\omega : \omega\omega\omega \prec \mathbb{F}[1, n]\}$ , the number of distinct cubes in  $\mathbb{F}[1, n]$ ;
- $D(n) := \#\{(\omega, p) : \omega_p\omega_{p+1}\omega_{p+2} \prec \mathbb{F}[1, n]\}$ , the number of repeated cubes in  $\mathbb{F}[1, n]$ .

The methods for counting the four functions have attracted some many authors, but known results are not rich. A.S.Fraenkel and J.Simpson gave the expression of  $A(f_m)$  and  $B(f_m)$  in 1999[8] and 2014[9]. In 2014, C.F.Du, H.Mousavi, L.Schaeffer and J.Shallit gave the expression of  $B(f_m)$  and  $D(f_m)$  by mechanical methods, see Theorem 58 and Theorem 59 in [7]. In this paper, we give the explicit expressions of  $A(n)$ ,  $B(f_m)$ ,  $C(n)$  and  $D(f_m)$ . Although we haven't get the explicit expressions of  $B(n)$  and  $D(n)$ , we give fast algorithms for counting  $B(n)$  and  $D(n)$  for all  $n$ .

The main tool of this paper is the “structure properties” of the sequence of return words in the Fibonacci sequence, which introduced and studied in [11], also see Property 2.2. The definition of return words is from F.Durand[6]. Let  $\omega$  be a factor of  $\mathbb{F}$ . For  $p \geq 1$ , let  $\omega_p = x_{i+1} \cdots x_{i+n}$  and  $\omega_{p+1} = x_{j+1} \cdots x_{j+n}$ . The factor  $x_{i+1} \cdots x_j$  is called the  $p$ -th return word of  $\omega$  and denoted by  $r_p(\omega)$ . The sequence  $\{r_p(\omega)\}_{p \geq 1}$  is called the sequences of the return words of factor  $\omega$ .

<sup>1</sup>School of Mathematics and Systems Science, Beihang University (BUAA), Beijing, 100191, P. R. China. E-mail address: huangyuke07@tsinghua.org.cn, hyg03ster@163.com.

<sup>2</sup>Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, P. R. China. E-mail address: wenzzy@tsinghua.edu.cn (Corresponding author).

By the “structure properties” (Property 2.2), we can determine the positions of all  $\omega_p$ . By the definition of square (resp. cube) and return word, we have

$$\omega_p \omega_{p+1} \prec \mathbb{F} \Leftrightarrow r_p(\omega) = \omega, \quad \omega_p \omega_{p+1} \omega_{p+2} \prec \mathbb{F} \Leftrightarrow r_p(\omega) = r_{p+1}(\omega) = \omega,$$

where the “=” means “have the same expressions”. By these relations, we can determine the positions of all squares and cubes, and then get  $A(n)$ ,  $B(n)$ ,  $C(n)$  and  $D(n)$ . But this method is complicated, another improved and fast method is used in this paper.

This paper is organized as follows. Section 2 present some basic notations and known results. Section 3 prove some basic properties of squares. We determine  $A(n)$  (distinct squares) in Section 4. Section 5 is devoted to establish the recursive structure of squares, then we determine  $B(n)$  (repeated squares) in Section 6. Similarly, we establish the recursive structure of cubes, then determine  $C(n)$  (distinct cubes) and  $D(n)$  (repeated cubes) in Section 7 to 10.

## 2 Preliminaries

Let  $\tau = x_1 \cdots x_n$  be a finite word (or  $\tau = x_1 x_2 \cdots$  be a sequence). For any  $i \leq j \leq n$ , define  $\tau[i, j] := x_i x_{i+1} \cdots x_{j-1} x_j$ . By convention, we denote  $\tau[i] := \tau[i, i] = x_i$  and  $\tau[i, i-1] = \varepsilon$ . Notation  $\nu \triangleright \omega$  means word  $\nu$  is a suffix of word  $\omega$ .

For  $m \geq -1$ , let  $\delta_m \in \{a, b\}$  be the last letter of  $F_m$ , then  $\delta_m = a$  iff  $m$  is even. The  $m$ -th singular word is defined as  $K_m = \delta_{m+1} F_m \delta_m^{-1} = \delta_{m+1} F_m [1, f_m - 1]$  for  $m \geq -1$ . By Property 2(9) in [18], all singular words are palindromes. Let  $Ker(\omega)$  be the maximal singular word occurring in factor  $\omega$ , called the kernel of  $\omega$ . Then by Theorem 1.9 in [11],  $Ker(\omega)$  occurs in  $\omega$  only once. Moreover

**Property 2.1** (Theorem 2.8 in [11]).  *$Ker(\omega_p) = Ker(\omega)_p$  for all  $\omega \in \mathbb{F}$  and  $p \geq 1$ .*

This means, let  $Ker(\omega) = K_m$ , then the maximal singular word occurring in  $\omega_p$  is just  $K_{m,p}$ . For instance,  $Ker(aba) = b$ ,  $(aba)_3 = \mathbb{F}[6, 8]$ ,  $(b)_3 = \mathbb{F}[7]$ , so  $Ker((aba)_3) = (b)_3$ ,  $(aba)_3 = a(b)_3 a$ .

**Property 2.2** (Theorem 2.11 in [11]). *For any factor  $\omega$ , the sequence of return words  $\{r_p(\omega)\}_{p \geq 1}$  is the Fibonacci sequence over the alphabet  $\{r_1(\omega), r_2(\omega)\}$ .*

Property 2.3 and 2.4 are useful in our proofs. Property 2.3 can be proved by induction. Since all singular words are palindromes, Property 2.4 holds by the cylinder structure of palindromes in [13].

**Property 2.3** (Lemma 2.2 in [11]). *For  $m \geq -1$ , (1)  $K_{m+3} = K_{m+1} K_m K_{m+1}$ .*

*(2)  $K_{m+2} = K_m K_{m+1} \delta_m^{-1} \delta_{m+1} = \delta_m^{-1} \delta_{m+1} K_{m+1} K_m$ .*

**Property 2.4.**  *$K_m \prec K_{m+3}[2, f_{m+3} - 1]$ ,  $K_{m+1} \not\prec K_{m+3}[2, f_{m+3} - 1]$ ,  $K_{m+2} \not\prec K_{m+3}[2, f_{m+3} - 1]$ .*

## 3 Basic properties of squares

By Definition 2.9 and Corollary 2.10 in [11], any factor  $\omega$  with kernel  $K_m$  can be expressed uniquely as  $\omega = K_{m+1}[i, f_{m+1}] K_m K_{m+1}[1, j] = K_{m+3}[i, f_{m+2} + j]$ , where  $2 \leq i \leq f_{m+1} + 1$  and  $0 \leq j \leq f_{m+1} - 1$ . By Property 2.1,  $\omega_p \omega_{p+1} \prec \mathbb{F}$  means

$$\omega_p \omega_{p+1} = K_{m+1}[i, f_{m+1}] \underbrace{K_{m,p} K_{m+1}[1, j] K_{m+1}[i, f_{m+1}]}_{r_p(K_m)} K_{m,p+1} K_{m+1}[1, j] \prec \mathbb{F}.$$

By Property 2.2,  $K_m$  has only two distinct return words  $r_1(K_m) = K_m K_{m+1}$  and  $r_2(K_m) = K_m K_{m-1}$ , so  $\omega_p \omega_{p+1} \prec \mathbb{F}$  has two cases as below, and in each case,  $|\omega| = |r_p(K_m)|$ .

**Case 1.**  $r_p(K_m) = r_1(K_m) = K_m K_{m+1}$ . Comparing the two expressions of  $r_p(K_m)$ , we have

$$K_m K_{m+1}[1, j] K_{m+1}[i, f_{m+1}] = K_m K_{m+1} \Rightarrow j = i - 1.$$

Comparing the two ranges of  $i$  that  $2 \leq i \leq f_{m+1} + 1$  and  $0 \leq j = i - 1 \leq f_{m+1} - 1$ , we have  $2 \leq i \leq f_{m+1}$  and  $m \geq 0$ . Moreover  $|\omega| = |r_1(K_m)| = f_{m+2}$  and

$$\begin{aligned}\omega\omega &= K_{m+1}[i, f_{m+1}]K_m K_{m+1}K_m K_{m+1}[1, i-1] \\ &= K_{m+2}[i, f_{m+2}]\underline{K_{m+1}}K_{m+2}[1, f_m + i - 1] = K_{m+4}[i, 2f_{m+2} + i - 1].\end{aligned}$$

The second and third equalities hold by Property 2.3.

Since  $K_{m+1} \prec \omega\omega \prec K_{m+4}[2, f_{m+4} - 1]$ , by Property 2.4,  $\text{Ker}(\omega\omega) = K_{m+1}$ .

**Case 2.**  $r_p(K_m) = r_2(K_m) = K_m K_{m-1}$ . Comparing the two expressions, we have

$$K_m K_{m+1}[1, j]K_{m+1}[i, f_{m+1}] = K_m K_{m-1} \Rightarrow j = i - f_m - 1.$$

So  $f_m + 1 \leq i \leq f_{m+1} + 1$  and  $m \geq -1$ . Moreover  $|\omega| = |r_2(K_m)| = f_{m+1}$  and

$$\begin{aligned}\omega\omega &= K_{m+1}[i, f_{m+1}]K_m K_{m-1}K_m K_{m+1}[1, i - f_m - 1] \\ &= K_{m+1}[i, f_{m+1}]\underline{K_{m+2}}K_{m+1}[1, i - f_m - 1] = K_{m+5}[f_{m+2} + i, f_{m+3} + f_{m+1} + i - 1].\end{aligned}$$

Since  $K_{m+2} \prec \omega\omega \prec K_{m+5}[2, f_{m+4} - 1]$ , by Property 2.4,  $\text{Ker}(\omega\omega) = K_{m+2}$ .

**Remark 3.1.** By the discussion above, we have: all squares in  $\mathbb{F}$  are of length  $2f_m$  for some  $m \geq 0$ ; for all  $m \geq 0$ , there exists a square of length  $2f_m$  in  $\mathbb{F}$ . This is a known result of P.S  bold[16].

**Property 3.2** (Property 4.1 in [13]).  $P(K_m, p) = pf_{m+1} + (\lfloor \phi p \rfloor + 1)f_m - 1$  for  $m \geq -1$ ,  $p \geq 1$ .

**Corollary 3.3** (Corollary 4.2 in [13]).  $P(a, p) = p + \lfloor \phi p \rfloor$ ,  $P(b, p) = 2p + \lfloor \phi p \rfloor$  for  $p \geq 1$ .

For  $m, p \geq 1$ , we define two sets below

$$\begin{cases} \langle 1, K_m, p \rangle := \{P(\omega\omega, p) : \text{Ker}(\omega\omega) = K_m, |\omega| = f_{m+1}, \omega\omega \prec \mathbb{F}\} \\ \langle 2, K_m, p \rangle := \{P(\omega\omega, p) : \text{Ker}(\omega\omega) = K_m, |\omega| = f_{m-1}, \omega\omega \prec \mathbb{F}\} \end{cases}$$

Obviously they correspond the two cases of squares respectively. By Property 3.2 we have

$$\begin{aligned}\langle 1, K_m, p \rangle &= \{P(\omega, p) : \omega = K_{m+1}[i, f_{m+1}]K_m K_{m+1}[1, f_{m-1} + i - 1], 2 \leq i \leq f_m\} \\ &= \{P(K_m, p) + f_{m-1} + i - 1, 2 \leq i \leq f_m\} \\ &= \{pf_{m+1} + \lfloor \phi p \rfloor f_m + f_{m+1}, \dots, pf_{m+1} + \lfloor \phi p \rfloor f_m + f_{m+2} - 2\}, \\ \langle 2, K_m, p \rangle &= \{P(\omega, p) : \omega = K_{m-1}[i, f_{m-1}]K_m K_{m-1}[1, i - f_{m-2} - 1], f_{m-2} + 1 \leq i \leq f_{m-1} + 1\} \\ &= \{P(K_m, p) + i - f_{m-2} - 1, f_{m-2} + 1 \leq i \leq f_{m-1} + 1\} \\ &= \{pf_{m+1} + \lfloor \phi p \rfloor f_m + f_m - 1, \dots, pf_{m+1} + \lfloor \phi p \rfloor f_m + 2f_{m-1} - 1\}.\end{aligned}$$

**Corollary 3.4.**  $\sharp\langle 1, K_m, p \rangle = f_m - 1$  and  $\sharp\langle 2, K_m, p \rangle = f_{m-3} + 1$  for  $m, p \geq 1$ .

## 4 The number of distinct squares in $\mathbb{F}[1, n]$

Denote  $a(n) := \sharp\{\omega : \omega\omega \triangleright \mathbb{F}[1, n], \omega\omega \not\prec \mathbb{F}[1, n-1]\}$ , obversely,  $A(n) = \sum_{i=1}^n a(i)$ . In order to count  $a(n)$ , we only need to consider  $\langle i, K_m, 1 \rangle$  where  $i = 1, 2$ .

**Property 4.1.**  $\langle 1, K_m, 1 \rangle = \{2f_{m+1}, \dots, f_{m+3} - 2\}$ ,  $\langle 2, K_m, 1 \rangle = \{f_{m+2} - 1, \dots, f_{m+1} + 2f_{m-1} - 1\}$ .

It is easy to see that sets  $\langle i, K_m, 1 \rangle$  are pairwise disjoint, and each set contains some consecutive integers. Therefore we get a chain

$$\langle 2, K_1, 1 \rangle, \langle 1, K_1, 1 \rangle, \dots, \langle 1, K_{m-1}, 1 \rangle, \langle 2, K_m, 1 \rangle, \langle 1, K_m, 1 \rangle, \langle 2, K_{m+1}, 1 \rangle, \dots$$

By this chain,  $a(n) = 1$  iff  $n \in \cup_{m \geq 1} (\langle 2, K_m, 1 \rangle \cup \langle 1, K_m, 1 \rangle)$ . The “ $\cup$ ” means pairwise disjoint union in this paper. Moreover, we have  $\langle 1, K_m, 1 \rangle \cup \langle 2, K_{m+1}, 1 \rangle = \{2f_{m+1}, \dots, f_{m+2} + 2f_m - 1\}$ .

**Property 4.2.**  $a(1) = a(2) = a(3) = 0$ ,  $a(4) = 1$  and for  $n \geq 5$

$$a(n) = 1 \text{ iff } n \in \cup_{m \geq 1} \{2f_{m+1}, \dots, f_{m+2} + 2f_m - 1\}.$$

One method for counting  $A(n)$  is by  $A(n) = \sum_{i=1}^n a(i)$ . By consider  $A(f_{m+2} + 2f_m - 1)$  for  $m \geq 1$ , we can give a fast algorithm of  $A(n)$  for all  $n \geq 1$ . Since  $\sum_{i=-1}^m f_i = f_{m+2} - 1$ ,

$$\begin{aligned} A(f_{m+2} + 2f_m - 1) &= a(4) + \sum_{i=1}^m \#\{2f_{i+1}, \dots, f_{i+2} + 2f_i - 1\} \\ &= 1 + \sum_{i=1}^m (f_i + f_{i-2}) = 1 + \sum_{i=-1}^m f_i - f_0 - f_{-1} + \sum_{i=-1}^{m-2} f_i = f_{m+2} + f_m - 3. \end{aligned}$$

**Theorem 4.3.** For all  $n \geq 1$ , let  $m$  satisfies  $2f_m \leq n < 2f_{m+1}$ ,

$$A(n) = \begin{cases} n - f_{m-1} - 2, & n \leq f_{m+1} + 2f_{m-1} - 1; \\ f_{m+1} + f_{m-1} - 3, & \text{otherwise.} \end{cases}$$

*Proof.* When  $f_{m+1} + 2f_{m-1} \leq n \leq 2f_{m+1} - 1$ ,  $a(n) = 0$ ,  $A(n) = A(f_{m+1} + 2f_{m-1} - 1) = f_{m+1} + f_{m-1} - 3$ .

When  $2f_m \leq n \leq f_{m+1} + 2f_{m-1} - 1$ ,  $a(n) = 1$ ,  $A(n) = A(2f_m - 1) + n - 2f_m + 1$ . Since  $A(2f_m - 1) = A(f_m + 2f_{m-2} - 1)$ , we have  $A(n) = n - f_{m-1} - 2$ . Thus the conclusion holds.  $\square$

**Remark 4.4.** Since  $2f_{m-2} \leq f_m \leq f_{m-1} + 2f_{m-3} - 1$  for  $m \geq 2$ , as a spacial case of Theorem 4.3,

$$A(f_m) = f_m - f_{m-3} - 2 = 2f_{m-2} - 2.$$

This is a known result of A.S.Fraenkel and J.Simpson, see Theorem 1 in [8].

## 5 The recursive structure of squares

In this section, we establish a recursive structure of squares. Using it, we will count the number of repeated squares in  $\mathbb{F}[1, n]$  (i.e.  $B(n)$ ) in Section 6. For  $m, p \geq 1$ , consider the vectors

$$\begin{aligned} \Gamma_{1,m,p} &:= [pf_{m+1} + [\phi p]f_m + f_{m+1} - 1, \langle 1, K_m, p \rangle] \\ &= [pf_{m+1} + [\phi p]f_m + f_{m+1} - 1, \dots, pf_{m+1} + [\phi p]f_m + f_{m+2} - 2]; \\ \Gamma_{2,m,p} &:= [\langle 2, K_m, p \rangle, pf_{m+1} + [\phi p]f_m + 2f_{m-1}, \dots, pf_{m+1} + [\phi p]f_m + f_{m+1} - 2] \\ &= [pf_{m+1} + [\phi p]f_m + f_m - 1, \dots, pf_{m+1} + [\phi p]f_m + f_{m+1} - 2]. \end{aligned}$$

Here vector  $[\langle i, K_m, p \rangle]$  means arrange all elements in set  $\langle i, K_m, p \rangle$ ,  $i = 1, 2$ .

Obversely, each  $\Gamma_{i,m,p}$  contains consecutive integers. The numbers of components in vectors  $\Gamma_{1,m,p}$  and  $\Gamma_{2,m,p}$  are  $f_m$  and  $f_{m-1}$  respectively. Moreover  $\max \Gamma_{2,m,p} + 1 = \min \Gamma_{1,m,p}$  for  $m, p \geq 1$ .

**Lemma 5.1** (Lemma 5.3 and 5.4 in [13]).  $[\phi(p + [\phi p] + 1)] = p$ ,  $[\phi(2p + [\phi p] + 1)] = p + [\phi p]$ .

**Property 5.2.**  $\Gamma_{1,m,p} = [\Gamma_{2,m-1,P(a,p)+1}, \Gamma_{1,m-1,P(a,p)+1}]$  for  $m \geq 2$ ,  $p \geq 1$ .

*Proof.* By Corollary 3.3,  $P(a, p) + 1 = p + [\phi p] + 1$ . By Lemma 5.1,  $[\phi(p + [\phi p] + 1)] = p$ .

$$\begin{aligned} \min \Gamma_{2,m-1,P(a,p)+1} &= (p + [\phi p] + 1)f_m + [\phi(p + [\phi p] + 1)]f_{m-1} + f_{m-1} - 1 \\ &= (p + [\phi p] + 1)f_m + pf_{m-1} + f_{m-1} - 1 = pf_{m+1} + [\phi p]f_m + f_{m+1} - 1 = \min \Gamma_{1,m,p}; \\ \max \Gamma_{1,m-1,P(a,p)+1} &= (p + [\phi p] + 1)f_m + [\phi(p + [\phi p] + 1)]f_{m-1} + f_{m+1} - 2 \\ &= (p + [\phi p] + 1)f_m + pf_{m-1} + f_{m+1} - 2 = pf_{m+1} + [\phi p]f_m + f_{m+2} - 2 = \max \Gamma_{1,m,p}. \end{aligned}$$

Since  $\max \Gamma_{2,m,p} + 1 = \min \Gamma_{1,m,p}$  for  $m, p \geq 1$ ,  $\max \Gamma_{2,m-1,P(a,p)+1} + 1 = \min \Gamma_{1,m-1,P(a,p)+1}$ . Thus the conclusion holds.  $\square$

By an analogous argument, we have

**Property 5.3.**  $\Gamma_{2,m,p} = [\Gamma_{2,m-2,P(b,p)+1}, \Gamma_{1,m-2,P(b,p)+1}]$  for  $m \geq 3, p \geq 1$ .

In Property 5.2 and 5.3, we establish the recursive relations for any  $\Gamma_{1,m,p}$  ( $m \geq 2$ ) and  $\Gamma_{2,m,p}$  ( $m \geq 3$ ). By the one-to-one correspondence between  $\Gamma_{i,m,p}$  and  $\langle i, K_m, p \rangle$ , we can define the recursive structure over  $\{\langle i, K_m, p \rangle \mid i = 1, 2; m, p \geq 1\}$  denoted by  $\mathcal{S}$ . Each  $\langle i, K_m, p \rangle$  is an element in  $\mathcal{S}$ . The recursive structure  $\mathcal{S}$  is a family of finite trees with root  $\langle i, K_m, 1 \rangle$  for all  $i = 1, 2, m \geq 1$ ; and with recursive relations:

$$\begin{cases} \tau_1 \langle 1, K_m, p \rangle = \langle 2, K_{m-1}, P(a, p) + 1 \rangle \cup \langle 1, K_{m-1}, P(a, p) + 1 \rangle & \text{for } m \geq 2; \\ \tau_2 \langle 2, K_m, p \rangle = \langle 2, K_{m-2}, P(b, p) + 1 \rangle \cup \langle 1, K_{m-2}, P(b, p) + 1 \rangle & \text{for } m \geq 3. \end{cases}$$

**Property 5.4.** Each  $\langle i, K_m, p \rangle$  belongs to the recursive structure  $\mathcal{S}$ ,  $i = 1, 2, m, p \geq 1$ .

*Proof.* Each element  $\langle i, K_m, 1 \rangle$  is root of a finite tree in  $\mathcal{S}$ . For  $m, p \geq 1$ ,

$$\begin{cases} \langle 1, K_m, P(a, p) + 1 \rangle \in \tau_1 \langle 1, K_{m+1}, p \rangle \\ \langle 1, K_m, P(b, p) + 1 \rangle \in \tau_2 \langle 2, K_{m+2}, p \rangle \end{cases} \quad \text{and} \quad \begin{cases} \langle 2, K_m, P(a, p) + 1 \rangle \in \tau_1 \langle 1, K_{m+1}, p \rangle \\ \langle 2, K_m, P(b, p) + 1 \rangle \in \tau_2 \langle 2, K_{m+2}, p \rangle \end{cases}$$

Since  $\mathbb{N} = \{1\} \cup \{P(a, p) + 1\} \cup \{P(b, p) + 1\}$ , the recursive structure  $\mathcal{S}$  contains all  $\langle i, K_m, p \rangle$ .  $\square$

On the other hand, by the recursive relations  $\tau_1$  and  $\tau_2$ , each element  $\langle i, K_m, p \rangle$  has a unique position in  $\mathcal{S}$ . By Property 5.2 and 5.3, the trees in  $\mathcal{S}$  are pairwise disjoint. Fig.1 and Fig.2 show the two finite trees in the recursive structure  $\mathcal{S}$  with roots  $\langle 1, K_5, 1 \rangle$  and  $\langle 2, K_5, 1 \rangle$  respectively.

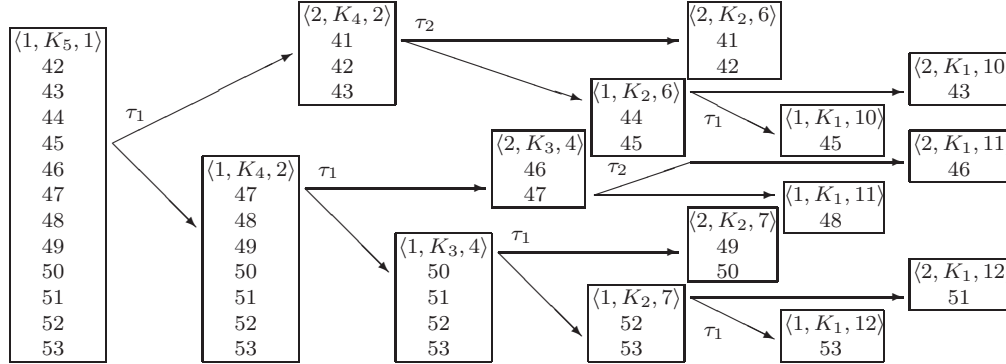


Fig.1: The finite tree in the recursive structure  $\mathcal{S}$  with root  $\langle 1, K_5, 1 \rangle$ .

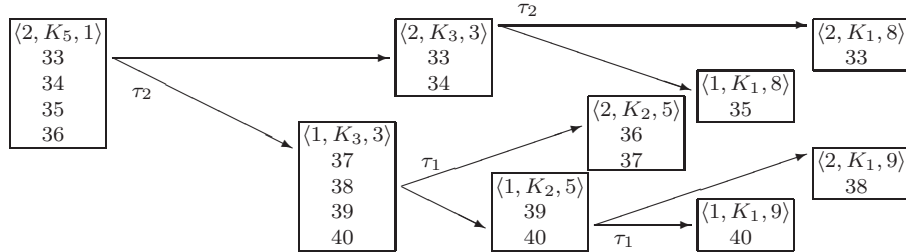


Fig.2: The finite tree in the recursive structure  $\mathcal{S}$  with root  $\langle 2, K_5, 1 \rangle$ .

By the recursive structure  $\mathcal{S}$ , we have the relation between the number of squares ending at position  $\Gamma_{1,m,p}[i] = pf_{m+1} + \lfloor \phi p \rfloor f_m + f_{m+1} + i - 1$  and  $\Gamma_{1,m,1}[i] = 2f_{m+1} + i - 1$ , see Property 5.5. Similarly, we have the relation between the number of squares ending at position  $\Gamma_{2,m,p}[i] = pf_{m+1} + \lfloor \phi p \rfloor f_m + f_m + i - 2$  and  $\Gamma_{2,m,1}[i] = f_{m+2} + i - 2$ , see Property 5.6.

**Property 5.5.** For  $1 \leq i \leq f_m - 1$ ,

$$\begin{aligned} & \{\omega : \omega\omega \triangleright \mathbb{F}[1, 2f_{m+1} + i - 1]\} \\ &= \{\omega : \omega\omega \triangleright \mathbb{F}[1, pf_{m+1} + \lfloor \phi p \rfloor f_m + f_{m+1} + i - 1], \text{Ker}(\omega) = K_j, 1 \leq j \leq m\}. \end{aligned}$$

**Property 5.6.** For  $1 \leq i \leq f_{m-3} + 1$ ,

$$\begin{aligned} & \{\omega : \omega\omega \triangleright \mathbb{F}[1, f_{m+2} + i - 2]\} \\ &= \{\omega : \omega\omega \triangleright \mathbb{F}[1, pf_{m+1} + \lfloor \phi p \rfloor f_m + f_m + i - 2], \text{Ker}(\omega) = K_j, 1 \leq j \leq m\}. \end{aligned}$$

For instance, taking  $m = 3$ ,  $p = 3$  and  $i = 2$  in the property above. All squares ending at position 13 are  $\{aabaab\}$ . All squares ending at position 34 are  $\{aabaab, abaabaababaabaab\}$ . Since  $\text{Ker}(aabaab) = aabaa = K_3$  and  $\text{Ker}(abaabaababaabaab) = aabaababaabaa = K_5$ , only  $\{aabaab\}$  is square with kernel  $K_j$ ,  $1 \leq j \leq 3$ . Fig.3 shows the relation:

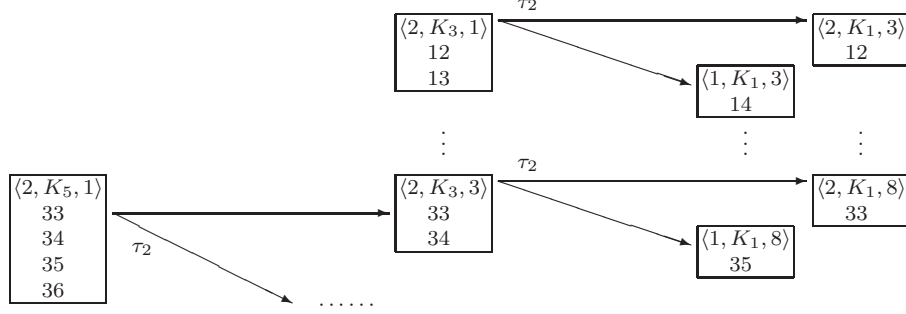


Fig.3: An example of the graph embedding in the recursive structure  $\mathcal{S}$ .

From Fig.3 we can see that: in the tree with root  $\langle 2, K_5, 1 \rangle$ , the branch from node  $\langle 2, K_3, 3 \rangle$  is the graph embedding of the tree with root  $\langle 2, K_3, 1 \rangle$ .

## 6 The number of repeated squares in $\mathbb{F}[1, n]$

Denote  $b(n) := \#\{(\omega, p) : \omega_p \omega_{p+1} \triangleright \mathbb{F}[1, n]\}$  the number of squares ending at position  $n$ . By the definition of  $\langle i, K_m, p \rangle$ ,  $b(n)$  is equal to the number of integer  $n$  occurs in the recursive structure  $\mathcal{S}$ . Thus we can calculate  $b(n)$  by the property below.

**Property 6.1.**  $b([1, 2, 3]) = [0, 0, 0]$ ,  $b(\Gamma_{2,1,1}) = b([4]) = [1]$ ,  $b(\Gamma_{1,1,1}) = b([5, 6]) = [0, 1]$ ,  $b(\Gamma_{2,2,1}) = b([7, 8]) = [1, 1]$ ,  $b(\Gamma_{1,2,1}) = b([9, 10, 11]) = [1, 1, 2]$ , for  $m \geq 3$ ,

$$\begin{cases} b(\Gamma_{1,m,1}) = [b(\Gamma_{2,m-1,1}), b(\Gamma_{1,m-1,1})] + [0, \underbrace{1, \dots, 1}_{f_m-1}]; \\ b(\Gamma_{2,m,1}) = [b(\Gamma_{2,m-2,1}), b(\Gamma_{1,m-2,1})] + [\underbrace{1, \dots, 1}_{f_{m-3}+1}, \underbrace{0, \dots, 0}_{f_{m-2}-1}]. \end{cases}$$

The first few values of  $b(n)$  are  $b([1, 2, 3]) = [0, 0, 0]$ ,  $b([4]) = [1]$ ,  $b([5, 6]) = [0, 1]$ ,  $b([7, 8]) = [1, 1]$ ,  $b([9, 10, 11]) = [1, 1, 2]$ ,  $b([12, 13, 14]) = [2, 1, 1]$ ,  $b([15, \dots, 19]) = [1, 2, 2, 2, 3]$ ,  $b([20, \dots, 24]) = [2, 2, 2, 1, 2]$ ,  $b([25, \dots, 32]) = [2, 2, 2, 2, 3, 3, 3, 4]$ .

For  $m \geq 3$ , the immediately corollaries are

$$\begin{cases} \sum b(\Gamma_{1,m,1}) = \sum b(\Gamma_{2,m-1,1}) + \sum b(\Gamma_{1,m-1,1}) + f_m - 1; \\ \sum b(\Gamma_{2,m,1}) = \sum b(\Gamma_{2,m-2,1}) + \sum b(\Gamma_{1,m-2,1}) + f_{m-3} + 1. \end{cases}$$

**Property 6.2.** For  $m \geq 1$ , (1)  $\sum b(\Gamma_{1,m,1}) = \frac{2m+5}{5}f_m + \frac{2m-6}{5}f_{m-2} - 1$ ,

(2)  $\sum b(\Gamma_{2,m,1}) = \frac{2m-2}{5}f_{m-1} + \frac{2m-3}{5}f_{m-3} + 1$ .

Since  $\Gamma_{1,m,1} = [2f_{m+1} - 1, \dots, f_{m+3} - 2]$  and  $\Gamma_{2,m,1} = [f_{m+2} - 1, \dots, 2f_{m+1} - 2]$ , we have  $B(f_{m+3}-2) = B(f_{m+2}-2) + \sum b(\Gamma_{2,m,1}) + \sum b(\Gamma_{1,m,1})$  and  $B(2f_{m+1}-2) = B(f_{m+2}-2) + \sum b(\Gamma_{2,m,1})$ . Thus by induction and Property 6.2,

**Property 6.3.** (1)  $B(f_{m+3}-2) = \frac{2m-4}{5}f_{m+3} + \frac{2m}{5}f_{m+1} + 4$  for  $m \geq -1$ .

(2)  $B(2f_{m+1}-2) = \frac{4m-11}{5}f_{m+1} + \frac{4m-3}{5}f_{m-1} + 5$  for  $m \geq 0$ .

Property 6.4 can be proved by induction and Property 6.1.

**Property 6.4.**  $b(f_m - 1) = \lfloor \frac{m-1}{2} \rfloor$ ,  $b(f_m) = \lfloor \frac{m}{2} - 1 \rfloor$ ,  $b(f_m - 1) + b(f_m) = m - 2$  for  $m \geq 2$ .

**Remark 6.5.** Since  $B(f_m) = B(f_m - 2) + b(f_m - 1) + b(f_m)$ ,  $m \geq 2$ . By Property 6.3 and 6.4,

$$B(f_m) = \frac{4}{5}(m+1)f_m - \frac{2}{5}(m+7)f_{m-1} - 4f_{m-2} + m + 2 = \frac{4m-16}{5}f_m - \frac{2m-6}{5}f_{m-1} + m + 2.$$

This is a known result of A.S.Fraenkel and J.Simpson[9].

Obversely we can calculate  $B(n)$  by  $B(n) = \sum_{i=4}^n b(i)$ . But when  $n$  is large, this method is complicated. Now we turn to give a fast algorithm. For any  $n \geq 4$ , let  $m$  such that  $f_m \leq n+1 < f_{m+1}$ . Since we already determine the expression of  $B(f_m - 2)$  and  $B(2f_{m-1} - 2)$  for  $m \geq 2$ , in order to give a fast algorithm of  $B(n)$ , we only need to calculate  $\sum_{i=f_m-1}^n b(i)$  or  $\sum_{i=2f_{m-1}-1}^n b(i)$ . One method is calculating  $b(n)$  by Property 6.1, the other method is using the corollaries as below.

**Corollary 6.6.** For  $n \geq 4$ , let  $m$  such that  $f_m \leq n+1 \leq 2f_{m-1} - 1$ , then  $m \geq 3$  and

$$\sum_{i=f_m-1}^n b(i) = \begin{cases} \sum_{i=f_{m-2}-1}^{n-f_{m-1}} b(i) + n - f_m + 2, & n+1 \leq f_m + f_{m-5} - 1; \\ \sum_{i=f_{m-2}+f_{m-5}-1}^{n-f_{m-1}} b(i) + \frac{2m-5}{5}f_{m-5} + \frac{2m-11}{5}f_{m-7} + 2, & \text{otherwise.} \end{cases}$$

*Proof.* By Property 6.1, when  $f_m \leq n+1 \leq f_m + f_{m-5} - 1$ ,

$$\sum_{i=f_m-1}^n b(i) = \sum_{i=f_{m-2}-1}^{n-f_{m-1}} [b(i) + 1] = \sum_{i=f_{m-2}-1}^{n-f_{m-1}} b(i) + n - f_m + 2.$$

When  $f_m + f_{m-5} \leq n+1 \leq 2f_{m-1} - 1$ ,  $\sum_{i=f_m-1}^n b(i) = \sum_{i=f_m-1}^{f_m+f_{m-5}-2} b(i) + \sum_{i=f_m+f_{m-5}-1}^n b(i)$ , where

$$\begin{cases} \sum_{i=f_m-1}^{f_m+f_{m-5}-2} b(i) = \sum_{i=f_{m-2}-1}^{f_{m-2}+f_{m-5}-2} [b(i) + 1] = \sum b(\Gamma_{2,m-4,1}) + f_{m-5} \\ \quad = \frac{2m-5}{5}f_{m-5} + \frac{2m-11}{5}f_{m-7} + 1; \\ \sum_{i=f_m+f_{m-5}-1}^n b(i) = \sum_{i=f_{m-2}+f_{m-5}-1}^{n-f_{m-1}} b(i) + 1. \end{cases}$$

Thus  $\sum_{i=f_m-1}^n b(i) = \sum_{i=f_{m-2}+f_{m-5}-1}^{n-f_{m-1}} b(i) + \frac{2m-5}{5}f_{m-5} + \frac{2m-11}{5}f_{m-7} + 2$ . The conclusion holds.  $\square$

**Corollary 6.7.** For  $n \geq 9$ , let  $m$  such that  $2f_{m-1} \leq n+1 \leq f_{m+1} - 1$ , then  $m \geq 4$  and

$$\sum_{i=2f_{m-1}-1}^n b(i) = \begin{cases} \sum_{i=f_{m-1}-1}^{n-f_{m-1}} b(i) + n - 2f_{m-1} + 1, & n+1 \leq f_m + f_{m-2} - 1; \\ \sum_{i=2f_{m-2}-1}^{n-f_{m-1}} b(i) + n - 2f_{m-1} + \frac{2m-8}{5}f_{m-4} + \frac{2m-9}{5}f_{m-6} + 2, & \text{otherwise.} \end{cases}$$

*Proof.* By Property 6.1, when  $2f_{m-1} \leq n+1 \leq 2f_{m-1} + f_{m-4} - 1 = f_m + f_{m-2} - 1$ ,

$$\sum_{i=2f_{m-1}-1}^n b(i) = \sum_{i=f_{m-1}-1}^{n-f_{m-1}} b(i) + \sum_{i=f_{m-1}}^{n-f_{m-1}} 1 = \sum_{i=f_{m-1}-1}^{n-f_{m-1}} b(i) + n - 2f_{m-1} + 1.$$



When  $f_m + f_{m-2} \leq n+1 \leq f_{m+1} - 1$ ,  $\sum_{i=2f_{m-1}-1}^n b(i) = \sum_{i=2f_{m-1}-1}^{f_m+f_{m-2}-2} b(i) + \sum_{i=f_m+f_{m-2}-1}^n b(i)$ , where

$$\left\{ \begin{array}{l} \sum_{i=2f_{m-1}-1}^{f_m+f_{m-2}-2} b(i) = \sum_{i=f_{m-1}-1}^{2f_{m-2}-2} b(i) + \sum_{i=f_{m-1}}^{2f_{m-2}-2} 1 = \sum_{i=f_{m-1}-1}^{2f_{m-2}-2} b(i) + f_{m-4} - 1 \\ = \sum b(\Gamma_{2,m-3,1}) + f_{m-4} - 1 = \frac{2m-3}{5}f_{m-4} + \frac{2m-9}{5}f_{m-6}; \\ \sum_{i=f_m+f_{m-2}-1}^n b(i) = \sum_{i=2f_{m-2}-1}^{n-f_{m-1}} [b(i) + 1] = \sum_{i=2f_{m-2}-1}^{n-f_{m-1}} b(i) + n - 2f_{m-1} - f_{m-4} + 2. \end{array} \right.$$

Thus  $\sum_{i=2f_{m-1}-1}^n b(i) = \sum_{i=2f_{m-2}-1}^{n-f_{m-1}} b(i) + n - 2f_{m-1} + \frac{2m-8}{5}f_{m-4} + \frac{2m-9}{5}f_{m-6} + 2$ . The conclusion holds.  $\square$

**Example.** One method to calculate  $\sum_{i=20}^{23} b(i)$  is by Property 6.1. Since  $b(\Gamma_{2,4,1}) = b([20, \dots, 24]) = [2, 2, 2, 1, 2]$ ,  $\sum_{i=20}^{23} b(i) = 7$ . The other method is using Corollary 6.6 and 6.7:

$$\sum_{i=20}^{23} b(i) = \sum_{i=f_4+f_1-1}^{23-f_5} b(i) + \frac{7}{5}f_1 + \frac{1}{5}f_{-1} + 2 = \sum_{i=9}^{10} b(i) + 5 = 7.$$

**Algorithm 6.8** (The number of repeated squares,  $B(n)$ ).

*Step 1.* For  $n \leq 3$ ,  $B(n) = 0$ ; for  $n \leq 4$ , find the  $m$  such that  $f_m \leq n+1 < f_{m+1}$ .

*Step 2.* Compare  $n$  with  $2f_{m-1} - 1$ .

(1) If  $n < 2f_{m-1} - 1$ , calculate  $B(f_m - 2)$  by Property 6.3; calculate  $\sum_{i=f_{m-1}}^n b(i)$  by Property 6.1 or by Corollary 6.6 and 6.7. Then  $B(n) = B(f_m - 2) + \sum_{i=f_{m-1}}^n b(i)$ .

(2) If  $n \geq 2f_{m-1} - 1$ , calculate  $B(2f_{m-1} - 2)$  by Property 6.3; calculate  $\sum_{i=2f_{m-1}-1}^n b(i)$  by Property 6.1 or by Corollary 6.6 and 6.7. Then  $B(n) = B(2f_{m-1} - 2) + \sum_{i=2f_{m-1}-1}^n b(i)$ .

**Remark 6.9.** When  $m$  is large (resp. small), Corollary 6.6 and 6.7 (resp. Property 6.1) is faster.

**Example.** We calculate  $B(23)$ . Since  $f_6 = 21 \leq 23 + 1 < f_7 = 34$ ,  $m = 6$ . Moreover  $23 < 2f_5 - 1$ .

By Property 6.3,  $B(f_6 - 2) = B(19) = \frac{2}{5}f_6 + \frac{6}{5}f_4 + 4 = 22$ . By Property 6.1 or by Corollary 6.6 and 6.7,  $\sum_{i=20}^{23} b(i) = 7$ . Thus  $B(23) = B(19) + \sum_{i=20}^{23} b(i) = 22 + 7 = 29$ .

## 7 Basic properties of cubes

Let  $\omega$  be a factor with kernel  $K_m$ , by an analogous argument as Section 3 and by Proposition 4.8 in [11],  $\omega_p \omega_{p+1} \omega_{p+2} \prec \mathbb{F}$  has only one case:  $r_p(K_m) = r_{p+1}(K_m) = r_1(K_m) = K_m K_{m+1}$ . In this case,  $|\omega| = f_{m+2}$ . Moreover  $2 \leq i \leq f_{m+1}$  and  $m \geq 0$ ,

$$\begin{aligned} \omega \omega \omega &= K_{m+1}[i, f_{m+1}] K_m K_{m+1} K_m K_{m+1} K_m K_{m+1} [1, i-1] \\ &= K_{m+2}[i, f_{m+2}] \underline{K_{m+3}} K_{m+2} [1, i + f_m - 1] = K_{m+6}[i + f_{m+3}, i + f_{m+5} + f_m - 1]. \end{aligned}$$

Since  $K_{m+3} \prec \omega \omega \omega \prec K_{m+6}[2, f_{m+6} - 1]$ , by Property 2.4,  $\text{Ker}(\omega \omega \omega) = K_{m+3}$ .

**Remark 7.1.** By the discussion above, we have: all cubes in  $\mathbb{F}$  are of length  $3f_m$  for some  $m \geq 2$ , and a cube of each such length occurs. This is Theorem 8 in J.Shallit et al[7].

For  $m \geq 3$  and  $p \geq 1$ , we define a set below:

$$\langle K_m, p \rangle := \{P(\omega \omega \omega, p) : \text{Ker}(\omega \omega \omega) = K_m, |\omega| = f_{m-1}, \omega \omega \omega \prec \mathbb{F}\}.$$

Obviously it contains all cubes. By Property 3.2 we have

$$\begin{aligned} \langle K_m, p \rangle &= \{P(\omega, p) : \omega = K_{m-1}[i, f_{m-1}] K_m K_{m-1} [1, i + f_{m-3} - 1], 2 \leq i \leq f_{m-2}\} \\ &= \{P(K_m, p) + f_{m-3} + i - 1, 2 \leq i \leq f_{m-2}\} \\ &= \{pf_{m+1} + \lfloor \phi p \rfloor f_m + 2f_{m-1}, \dots, pf_{m+1} + \lfloor \phi p \rfloor f_m + f_{m+1} - 2\}. \end{aligned}$$

**Corollary 7.2.**  $\sharp \langle K_m, p \rangle = f_{m-2} - 1$  for  $m \geq 3$ ,  $p \geq 1$ .



## 8 The number of distinct cubes in $\mathbb{F}[1, n]$

Denote  $c(n) := \#\{\omega : \omega\omega\omega \triangleright \mathbb{F}[1, n], \omega\omega\omega \not\triangleright \mathbb{F}[1, n-1]\}$ . Obversely,  $C(n) = \sum_{i=1}^n c(i)$ .

**Property 8.1.**  $\langle K_m, 1 \rangle = \{f_{m+1} + 2f_{m-1}, \dots, 2f_{m+1} - 2\}$  for  $m \geq 3$ .

Sets  $\langle K_m, 1 \rangle$  are pairwise disjoint, and each set contains some consecutive integers. We get a chain  $\langle K_3, 1 \rangle = \{14\}, \langle K_4, 1 \rangle = \{23, 24\}, \dots, \langle K_m, 1 \rangle, \dots$ . So  $c(n) = 1$  iff  $n \in \cup_{m \geq 3} \langle K_m, 1 \rangle$ . Thus

**Property 8.2.**  $c(n) = 1$  iff  $n \in \cup_{m \geq 3} \{f_{m+1} + 2f_{m-1}, \dots, 2f_{m+1} - 2\}$ .

By consider  $C(2f_{m+1} - 2)$  for  $m \geq 3$ , we can give a fast algorithm of  $C(n)$  for all  $n \geq 1$ . Since  $\sum_{i=1}^m f_i = f_{m+2} - 1$ ,  $C(2f_{m+1} - 2) = \sum_{i=3}^m \#\langle K_i, 1 \rangle = \sum_{i=3}^m (f_{i-2} - 1) = f_m - m - 1$ .

**Theorem 8.3.** For  $n < 14$ ,  $C(n) = 0$ ; for  $n \geq 14$ , let  $m$  s.t.  $f_{m+1} + 2f_{m-1} \leq n < f_{m+2} + 2f_m - 1$ , then  $m \geq 3$  and

$$C(n) = \begin{cases} n - f_{m+1} - f_{m-1} - m + 1, & n \leq 2f_{m+1} - 2; \\ f_m - m - 1, & \text{otherwise.} \end{cases}$$

*Proof.* When  $2f_{m+1} - 1 \leq n \leq f_{m+2} + 2f_m - 1$ ,  $c(n) = 0$ ,  $C(n) = C(2f_{m+1} - 2) = f_m - m - 1$ .

When  $f_{m+1} + 2f_{m-1} \leq n \leq 2f_{m+1} - 2$ ,  $c(n) = 1$ ,

$$C(n) = C(f_{m+1} + 2f_{m-1} - 1) + n - f_{m+1} - 2f_{m-1} + 1.$$

Since  $C(f_{m+1} + 2f_{m-1} - 1) = C(2f_m - 2) = f_{m-1} - m$ ,  $C(n) = n - f_{m+1} - f_{m-1} - m + 1$ . Thus the conclusion holds.  $\square$

Since  $2f_{m-2} - 1 \leq f_m \leq f_{m-1} + 2f_{m-3} - 1$  for  $m \geq 6$ , we have

**Theorem 8.4.**  $C(f_m) = 0$  for  $m \leq 5$ ,  $C(f_m) = f_{m-3} - m + 2$  for  $m \geq 6$ .

## 9 The recursive structure of cubes

In this section, we establish a recursive structure of cubes. Using it, we will count the number of repeated cubes in  $\mathbb{F}[1, n]$  (i.e.  $D(n)$ ) in Section 10.

**Property 9.1.** For  $m \geq 5$ ,  $\min\langle K_m, p \rangle - 2 = \max\langle K_{m-2}, P(b, p) + 1 \rangle$ .

*Proof.* Since  $P(b, p) = 2p + \lfloor \phi p \rfloor$ ,  $\lfloor \phi(2p + \lfloor \phi p \rfloor + 1) \rfloor = p + \lfloor \phi p \rfloor$ , for  $m \geq 3$ ,  $f_{m-1} + f_{m-4} = 2f_{m-2}$ ,

$$\begin{aligned} \max\langle K_{m-2}, P(b, p) + 1 \rangle + 2 &= (2p + \lfloor \phi p \rfloor + 1)f_{m-1} + \lfloor \phi(2p + \lfloor \phi p \rfloor + 1) \rfloor f_{m-2} + f_{m-1} \\ &= (2p + \lfloor \phi p \rfloor + 1)f_{m-1} + (p + \lfloor \phi p \rfloor)f_{m-2} + f_{m-1} = pf_{m+1} + \lfloor \phi p \rfloor f_m + 2f_{m-1} = \min\langle K_m, p \rangle. \end{aligned}$$

This means  $\max\langle K_{m-2}, P(b, p) + 1 \rangle + 2 = \min\langle K_m, p \rangle$ , so the conclusion holds.  $\square$

By an analogous argument, we have

**Property 9.2.** For  $m \geq 4$ ,  $\max\langle K_m, p \rangle + f_{m-4} + 2 = \min\langle K_{m-1}, P(a, p) + 1 \rangle$ .

In Property 9.1 and 9.2, we establish the recursive relations for any  $\langle K_m, p \rangle$ ,  $m \geq 3$ . Thus we can define the recursive structure over  $\{\langle K_m, p \rangle \mid m \geq 3, p \geq 1\}$  denoted by  $\mathcal{C}$ . Each  $\langle K_m, p \rangle$  is an element in  $\mathcal{C}$ . The recursive structure  $\mathcal{C}$  is a family of finite trees with roots  $\langle K_m, 1 \rangle$  for all  $m \geq 3$ ; and with recursive relations:

$$\begin{cases} \tau_3\langle K_m, p \rangle = \langle K_{m-2}, P(b, p) + 1 \rangle \cup \langle K_{m-1}, P(a, p) + 1 \rangle \text{ for } m \geq 5; \\ \tau_4\langle K_4, p \rangle = \langle K_{m-1}, P(a, p) + 1 \rangle. \end{cases}$$

Since  $\max\langle K_{m-2}, P(b, p) + 1 \rangle < \min\langle K_{m-1}, P(a, p) + 1 \rangle$ , the “ $\cup$ ” is a disjoint union.

**Property 9.3.** Each  $\langle K_m, p \rangle$  belongs to the recursive structure  $\mathcal{C}$ , for  $m \geq 3$  and  $p \geq 1$ .

*Proof.* Each element  $\langle K_m, 1 \rangle$  is root of a finite tree in  $\mathcal{C}$ . For  $p \geq 1$ ,

$$\begin{cases} \langle K_m, P(a, p) + 1 \rangle \in \tau_3 \langle K_{m+1}, p \rangle \ (m \geq 4) \text{ and } \langle K_3, P(a, p) + 1 \rangle \in \tau_4 \langle K_4, p \rangle; \\ \langle K_m, P(b, p) + 1 \rangle \in \tau_3 \langle K_{m+2}, p \rangle \ (m \geq 3). \end{cases}$$

Since  $\mathbb{N} = \{1\} \cup \{P(a, p) + 1\} \cup \{P(b, p) + 1\}$ , the recursive structure  $\mathcal{C}$  contains all  $\langle K_m, p \rangle$ .  $\square$

On the other hand, by the recursive relations  $\tau_3$  and  $\tau_4$ , each element  $\langle K_m, p \rangle$  has a unique position in  $\mathcal{C}$ . Fig.4 show the finite tree in the recursive structure  $\mathcal{C}$  with root  $\langle K_6, 1 \rangle$ .

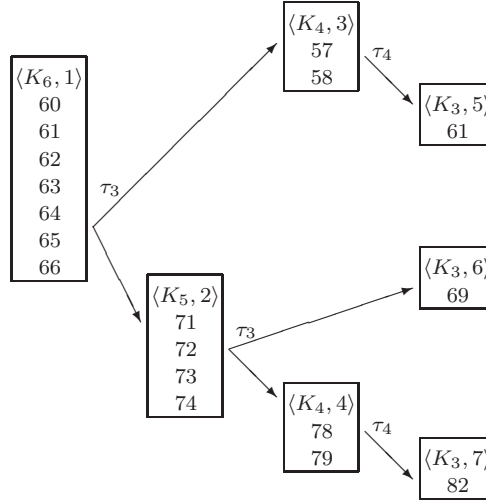


Fig.4: The finite tree in the recursive structure  $\mathcal{C}$  with root  $\langle K_6, 1 \rangle$ .

**Lemma 9.4.** For  $m \geq 1$ , (1)  $P(a, f_m - 1) = f_{m+1} - 2$ ,  $\lfloor \phi(f_m - 1) \rfloor = f_{m-1} - 1$ .

(2)  $\lfloor \phi f_m \rfloor = f_{m-1}$  if  $m$  is odd;  $\lfloor \phi f_m \rfloor = f_{m-1} - 1$  if  $m$  is even. (3)  $P(b, f_{2m}) = f_{2m+2} - 1$ .

*Proof.* Denote by  $|\omega|_a$  (resp.  $|\omega|_b$ ) the number of letter  $a$  (resp.  $b$ ) occurring in  $\omega$ .

(1) Since  $|F_{m+1}|_a = f_m$ ,  $aba \triangleright F_{2m}$  and  $aab \triangleright F_{2m+1}$ , we have  $P(a, f_m - 1) = f_{m+1} - 2$ . On the other hand, by Corollary 3.3,  $P(a, f_m - 1) = f_m - 1 + \lfloor \phi(f_m - 1) \rfloor$ . Comparing the two expressions of  $P(a, f_m - 1)$ , we have  $\lfloor \phi(f_m - 1) \rfloor = f_{m-1} - 1$  for  $m \geq 1$ .

(2) By Corollary 3.3,  $P(a, f_m) = f_m + \lfloor \phi f_m \rfloor$ . By the analogous argument in (1), we have: when  $m$  is odd,  $P(a, f_m) = f_{m+1}$ , then  $P(a, f_m) = f_{m+1} = f_m + \lfloor \phi f_m \rfloor \Rightarrow \lfloor \phi f_m \rfloor = f_{m-1}$ ; when  $m$  is even,  $P(a, f_m) = f_{m+1} - 1$ , then  $P(a, f_m) = f_{m+1} - 1 = f_m + \lfloor \phi f_m \rfloor \Rightarrow \lfloor \phi f_m \rfloor = f_{m-1} - 1$ .

(3) Since  $|F_m|_b = f_{m-2}$ ,  $aba \triangleright F_{2m}$ , we have  $P(b, f_{2m}) = f_{2m+2} - 1$  for  $m \geq 1$ .  $\square$

**Lemma 9.5.**  $f_m f_k + f_{m-1} f_{k-1} = f_{m+k+1}$  for  $m, k \geq -1$ .

*Proof.* Since  $f_m f_k + f_{m-1} f_{k-1} = f_m(f_{k-1} + f_{k-2}) + f_{m-1} f_{k-1} = f_m f_{k-2} + (f_m + f_{m-1}) f_{k-1}$ , using it repeatedly,  $f_m f_k + f_{m-1} f_{k-1} = f_m f_{k-2} + f_{m+1} f_{k-1} = \dots = f_{m+k-1} f_{-1} + f_{m+k} f_0 = f_{m+k+1}$ .  $\square$

For  $m \geq 3$ , we define the vectors  $\Gamma_m := [f_{m+2} - 1, \dots, f_{m+3} - 2]$ , then

**Property 9.6.** The finite tree with root  $\langle K_m, 1 \rangle$  belongs to  $\Gamma_m$  for  $m \geq 3$ .

*Proof.* (1) Since  $P(a, f_m - 1) = f_{m+1} - 2$ , the maximal of the recursive structure from  $\langle K_m, 1 \rangle$  is

$$\max\{\max\langle K_m, 1 \rangle, \max\langle K_{m-1}, f_2 - 1 \rangle, \max\langle K_{m-2}, f_3 - 1 \rangle, \dots, \max\langle K_3, f_{m-2} - 1 \rangle\}$$

By Property 9.2,  $\max\langle K_m, p \rangle < \min\langle K_{m-1}, P(a, p) + 1 \rangle$ , so  $\max\langle K_{m-i}, f_{i+1} - 1 \rangle$  is strictly increasing for  $0 \leq i \leq m - 3$ . Thus the maximal integer in the tree is  $\max\langle K_3, f_{m-2} - 1 \rangle$ .

$$\begin{aligned} \max\langle K_3, f_{m-2} - 1 \rangle &= (f_{m-2} - 1)f_4 + \lfloor \phi(f_{m-2} - 1) \rfloor f_3 + f_4 - 2 \\ &= (f_{m-2} - 1)f_4 + (f_{m-3} - 1)f_3 + 6 = f_{m-2}f_4 + f_{m-3}f_3 - 7 = f_{m+3} - 7 < \max \Gamma_m. \end{aligned}$$

(2) Similarly, since  $P(b, f_{2m}) = f_{2m+2} - 1$ ,  $\min\langle K_{m-2i}, f_{2i} \rangle$  is strictly decreasing for  $0 \leq i \leq \lfloor \frac{m-4}{2} \rfloor$ . So the minimal integer in the tree is

$$\min\{\min\langle K_m, 1 \rangle, \min\langle K_{m-2}, f_2 \rangle, \min\langle K_{m-4}, f_4 \rangle, \dots\} = \begin{cases} \min\langle K_4, f_{m-4} \rangle & \text{if } m \text{ is even;} \\ \min\langle K_3, f_{m-3} \rangle & \text{if } m \text{ is odd.} \end{cases}$$

When  $m$  is even,  $\min\langle K_4, f_{m-4} \rangle = f_{m-4}f_5 + \lfloor \phi f_{m-4} \rfloor f_4 + 2f_3$ ,  $\lfloor \phi f_{m-4} \rfloor = f_{m-5} - 1$ , so

$$\min\langle K_4, f_{m-4} \rangle = f_{m-4}f_5 + (f_{m-5} - 1)f_4 + 2f_3 = f_{m+2} + 2 > \min \Gamma_m.$$

When  $m$  is odd,  $\min\langle K_3, f_{m-3} \rangle = f_{m-3}f_4 + \lfloor \phi f_{m-3} \rfloor f_3 + 2f_2$ ,  $\lfloor \phi f_{m-3} \rfloor = f_{m-4} - 1$ , so

$$\min\langle K_3, f_{m-3} \rangle = f_{m-3}f_4 + (f_{m-4} - 1)f_3 + 2f_2 = f_{m+2} + 1 > \min \Gamma_m.$$

In each case, the minimal integer in the tree is larger than  $\min \Gamma_m$ , so the conclusion holds.  $\square$

By Property 9.6 and the definition of  $\Gamma_m$ , the finite trees in recursive structure  $\mathcal{C}$  with different roots  $\langle K_m, 1 \rangle$  are disjoint.

## 10 The number of repeated cubes in $\mathbb{F}[1, n]$

Denote  $d(n) := \sharp\{(\omega, p) : \omega_p \omega_{p+1} \omega_{p+2} \triangleright \mathbb{F}[1, n]\}$ , the number of cubes ending at position  $n$ . Obversely,  $D(n) = \sum_{i=1}^n d(i)$ . By the definition of  $\langle K_m, p \rangle$ ,  $d(n)$  is equal to the number of integer  $n$  occurs in the recursive structure  $\mathcal{C}$ . Thus we can calculate  $d(n)$  by the property below.

**Property 10.1.** For  $m \geq 3$ ,

$$\begin{aligned} & d([f_{m+4} - 1, \dots, f_{m+5} - 2]) \\ &= d([f_{m+2} - 1, \dots, f_{m+3} - 2, f_{m+3} - 1, \dots, f_{m+4} - 2]) + \underbrace{[0, \dots, 0]}_{f_{m-1}+1}, \underbrace{[1, \dots, 1]}_{f_{m-1}}, \underbrace{[0, \dots, 0]}_{f_{m+2}}. \end{aligned}$$

The first few values of  $d(n)$  are  $d([f_5 - 1, \dots, f_6 - 2]) = [d(12), \dots, d(19)] = [0, 0, 1, 0, 0, 0, 0, 0]$ ,  
 $d([f_6 - 1, \dots, f_7 - 2]) = [d(20), \dots, d(32)] = [0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0]$ ,  
 $d([f_7 - 1, \dots, f_8 - 2]) = [d(33), \dots, d(53)] = [0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0]$ .

By Property 10.1,  $\sum d(\Gamma_{m+2}) = \sum d(\Gamma_m) + \sum d(\Gamma_{m+1}) + f_{m-2} - 1$ . By induction, we have

**Lemma 10.2.**  $\sum_{f_{m+2}-1}^{f_{m+3}-2} d(n) = \frac{m-5}{5}f_m + \frac{m+2}{5}f_{m-2} + 1$  for  $m \geq 3$ .

By the definition of  $D(n)$ ,  $D(f_{m+1} - 2) = D(f_m - 2) + \sum_{f_m-1}^{f_{m+1}-2} d(n)$ . By induction, we have

**Property 10.3.**  $D(f_m - 2) = \frac{m-11}{5}f_{m-1} + \frac{m+1}{5}f_{m-3} + m + 1$  for  $m \geq 6$ .

By Property 10.1, we get  $d(f_m - 1) = d(f_m) = 0$  easily by induction, thus  $D(f_m) = D(f_m - 2)$ .

**Theorem 10.4.**  $D(f_m) = \frac{m-11}{5}f_{m-1} + \frac{m+1}{5}f_{m-3} + m + 1$  for  $m \geq 6$ .

**Remark 10.5.** Theorem 59 in [7] shows the number of cube occurrences in  $F_m$  as

$$D(f_m) = [d_1(m+2) + d_2]\alpha^{m+2} + [d_3(m+2) + d_4]\beta^{m+2} + m + 1.$$

where  $\alpha = \frac{1+\sqrt{5}}{2}$ ,  $\beta = \frac{1-\sqrt{5}}{2}$ ,  $d_1 = \frac{3-\sqrt{5}}{10}$ ,  $d_2 = \frac{17}{50}\sqrt{5} - \frac{3}{2}$ ,  $d_3 = \frac{3+\sqrt{5}}{10}$ ,  $d_4 = -\frac{17}{50}\sqrt{5} - \frac{3}{2}$ . Since

$$f_m = \frac{\alpha^{m+2} - \beta^{m+2}}{\alpha - \beta}, \quad \alpha\beta = -1, \quad \frac{1}{\alpha} = \frac{\sqrt{5}-1}{2}, \quad \frac{1}{\beta} = \frac{-1-\sqrt{5}}{2}, \quad \left(\frac{1}{\alpha}\right)^3 = \sqrt{5} - 2, \quad \left(\frac{1}{\beta}\right)^3 = -\sqrt{5} - 2,$$

we can prove the two expressions are same. By our expression in Theorem 10.4,

$$\begin{aligned}
D(f_m) - m - 1 &= \frac{m-11}{5} \times \frac{\alpha^{m+1}-\beta^{m+1}}{\alpha-\beta} + \frac{m+1}{5} \times \frac{\alpha^{m-1}-\beta^{m-1}}{\alpha-\beta} \\
&= \frac{m-11}{5\sqrt{5}} \times \left( \frac{\sqrt{5}-1}{2} \alpha^{m+2} + \frac{1+\sqrt{5}}{2} \beta^{m+2} \right) + \frac{m+1}{5\sqrt{5}} \times ((\sqrt{5}-2)\alpha^{m+2} + (\sqrt{5}+2)\beta^{m+2}) \\
&= \left[ \frac{m-11}{5\sqrt{5}} \times \frac{\sqrt{5}-1}{2} + \frac{m+1}{5\sqrt{5}}(\sqrt{5}-2) \right] \alpha^{m+2} + \left[ \frac{m-11}{5\sqrt{5}} \times \frac{1+\sqrt{5}}{2} + \frac{m+1}{5\sqrt{5}}(\sqrt{5}+2) \right] \beta^{m+2} \\
&= \left[ \frac{3-\sqrt{5}}{10} m + \frac{7\sqrt{5}-45}{50} \right] \alpha^{m+2} + \left[ \frac{3+\sqrt{5}}{10} m + \frac{-7\sqrt{5}-45}{50} \right] \beta^{m+2}.
\end{aligned}$$

By J.shallit's expression in [7],

$$D(f_m) - m - 1 = \left[ \frac{3-\sqrt{5}}{10} m + \frac{3-\sqrt{5}}{5} + \frac{17}{50} \sqrt{5} - \frac{3}{2} \right] \alpha^{m+2} + \left[ \frac{3+\sqrt{5}}{10} m + \frac{3+\sqrt{5}}{5} - \frac{17}{50} \sqrt{5} - \frac{3}{2} \right] \beta^{m+2}.$$

Comparing the coefficients of  $m\alpha^{m+2}$ ,  $\alpha^{m+2}$ ,  $m\beta^{m+2}$  and  $\beta^{m+2}$ , we have the two expressions are same.

For any  $n \geq 12$ , let  $m$  such that  $f_m \leq n+1 < f_{m+1}$ . Since we already determine the expression of  $D(f_m - 2)$ , in order to give a fast algorithm of  $D(n)$ , we only need to calculate  $\sum_{i=f_m-1}^n d(i)$ . One method is calculating  $d(n)$  by Property 10.1, the other method is using the corollaries as below.

**Corollary 10.6.** For  $n \geq 12$ , let  $m$  such that  $f_m \leq n+1 < f_{m+1}$ , then  $m \geq 5$  and

$$\sum_{i=f_m-1}^n d(i) = \begin{cases} \sum_{i=f_m-2-1}^{n-f_m-1} d(i), & f_m \leq n+1 \leq f_m + f_{m-5}; \\ \sum_{i=f_m-2+1}^{n-f_m-1} d(i) + n - f_m - f_{m-5} + 1, & f_m + f_{m-5} + 1 \leq n+1 \leq f_m + f_{m-3} - 1; \\ \sum_{i=f_{m-1}-1}^{n-f_m-1} d(i) + \frac{m-4}{5} f_{m-4} + \frac{m-2}{5} f_{m-6}, & f_m + f_{m-3} \leq n+1 < f_{m+1}. \end{cases}$$

*Proof.* By Property 10.1, when  $f_m \leq n+1 \leq f_m + f_{m-5}$ ,  $\sum_{i=f_m-1}^n d(i) = \sum_{i=f_m-2-1}^{n-f_m-1} d(i)$ .

When  $f_m + f_{m-5} + 1 \leq n+1 \leq f_m + f_{m-3} - 1$ ,

$$\begin{aligned}
\sum_{i=f_m-1}^n d(i) &= \sum_{i=f_m-1}^{f_m+f_{m-5}-1} d(i) + \sum_{i=f_m+f_{m-5}}^n d(i) = \sum_{i=f_m-2+1}^{n-f_m-1} d(i) + \sum_{i=f_m-2+f_{m-5}}^{n-f_m-1} 1 \\
&= \sum_{i=f_m-2+1}^{n-f_m-1} d(i) + n - f_m - f_{m-5} + 1.
\end{aligned}$$

When  $f_m + f_{m-3} \leq n+1 < f_{m+1}$ ,

$$\begin{aligned}
\sum_{i=f_m-1}^n d(i) &= \sum_{i=f_m-1}^{f_m+f_{m-5}-1} d(i) + \sum_{i=f_m+f_{m-5}}^{f_m+f_{m-3}-2} d(i) + \sum_{i=f_m+f_{m-3}-1}^n d(i) \\
&= \sum_{i=f_m-2-1}^{f_{m-1}-2} d(i) + f_{m-4} - 1 + \sum_{i=f_m-1-1}^{n-f_m-1} d(i) = \sum_{i=f_{m-1}-1}^{n-f_m-1} d(i) + \frac{m-4}{5} f_{m-4} + \frac{m-2}{5} f_{m-6}.
\end{aligned}$$

So the conclusion holds.  $\square$

**Example.** Since  $[d(33), \dots, d(53)] = [0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0]$ , using Property 10.1, we have  $\sum_{i=33}^{48} d(i) = 8$ . The other method is using Corollary 10.6. Since  $f_7 + f_4 = 42 \leq n+1 = 49 < f_8 = 55$ ,  $\sum_{i=33}^{48} d(i) = \sum_{i=f_6-1}^{48-f_6} d(i) + \frac{3}{5} f_3 + \frac{5}{5} f_1 = \sum_{i=20}^{27} d(i) + 5 = \sum_{i=12}^{14} d(i) + 7 = 8$ .

**Algorithm 10.7** (The number of repeated cubes occurrences,  $D(n)$ ).

- Step 1. For  $n \leq 11$ ,  $D(n) = 0$ ; for  $n \leq 12$ , find the  $m$  such that  $f_m \leq n+1 < f_{m+1}$ .
- Step 2. Calculate  $D(f_m - 2)$  by Property 10.3.
- Step 3. Calculate  $\sum_{i=f_m-1}^n d(i)$  by Property 10.1 or by Corollary 10.6.
- Step 4.  $D(n) = D(f_m - 2) + \sum_{i=f_m-1}^n d(i)$ .

**Example.** We calculate  $D(48)$ . Since  $f_7 = 34 \leq 48 + 1 < f_8 = 55$ ,  $m = 7$ .

By Property 10.3,  $D(32) = D(f_7 - 2) = \frac{-4}{5}f_6 + \frac{8}{5}f_4 + 7 + 1 = 4$ .

By Property 10.1 or by Corollary 10.6,  $\sum_{i=33}^{48} d(i) = 8$ . Thus  $D(48) = D(32) + \sum_{i=33}^{48} d(i) = 12$ .

## Acknowledgments

The research is supported by the Grant NSF No.11431007, No.11271223 and No.11371210.

## References

- [1] J.M.Allouche, J.Shallit. Automatic sequences: Theory, applications, generalizations. Cambridge University Press, Cambridge, 2003.
- [2] J.Berstel. Recent results in Sturmian words, in J.Dassow, A.Salomaa (Eds.), Developments in Language Theory, World Scientific, Singapore. (1966) 13-24.
- [3] J.Berstel. Mot de Fibonacci, Séminaire d'informatique théorique, L.I.T.P., Paris, 1980/1981, 57-78.
- [4] W.-F.Chuan, H.-L.Ho. Locating factors of the infinite Fibonacci word, Theoretical Computer Science. 349 (2005) 429-442.
- [5] W.-T.Cao, Z.-Y.Wen. Some properties of the factors of Sturmian sequences, Theoretical Computer Science. 304 (2003) 365-385.
- [6] F.Durand. A characterization of substitutive sequences using return words, Discrete Math. 179 (1998) 89-101.
- [7] C.-F.Du, H.Mousavi, L.Schaeffer, J.Shallit. Decision Algorithms for Fibonacci-Automatic Words, with Applications to Pattern Avoidance. Eprint Arxiv, 2014.
- [8] A.S.Fraenkel, J.Simpson. The exact number of squares in Fibonacci words, Theoretical Computer Science. 218 (1999) 95-106.
- [9] A.S.Fraenkel, J.Simpson. Corrigendum to "The exact number of squares in Fibonacci words", Theoretical Computer Science. 547 (2014) 122.
- [10] A.Glen. On Sturmian and Episturmian Words, and Related Topics, PhD thesis, The University of Adelaide, Australia. 2006.
- [11] Y.-K.Huang, Z.-Y.Wen. The sequence of return words of the Fibonacci sequence, Theoretical Computer Science. 593 (2015) 106-116.
- [12] Y.-K.Huang, Z.-Y.Wen. Kernel words and gap sequence of the Tribonacci sequence, Acta Mathematica Scientia (Series B). 36.1 (2016) 173-194.
- [13] Y.-K.Huang, Z.-Y.Wen. The structure of palindromes in the Fibonacci sequence. arXiv: 1601.04391.
- [14] M.Lothaire. Combinatorics on words, in: Encyclopedia of Mathematics and its applications, Vol.17, Addison-Wesley, Reading, MA, 1983.
- [15] M.Lothaire. Algebraic combinatorics on words, Cambridge Univ. Press, Cambridge, 2002.
- [16] P.Séébold. Propriétés combinatoires des mots infinis engendrés par certains morphismes (Thèse de 3<sup>e</sup> cycle). PhD thesis, Université P. et M. Curie, Institut de Programmation, Paris, 1985.

- [17] B.Tan, Z.-X.Wen. Invertible substitutions and Sturmian sequences. *European Journal of Combinatorics*, 24.8 (2003) 983-1002.
- [18] Z.-X.Wen, Z.-Y.Wen. Some properties of the singular words of the Fibonacci word, *European J. Combin.* 15 (1994) 587-598.